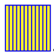
9.4 Computing $\mathbf{u}^*$ via orthogonalisation (QR decomposition)

QR-decompostion presents a much more stable solution to the normal equation $\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{u} = \mathbf{A}^{\mathrm{T}}\mathbf{b}$.

The Orthogonalisation of matrix $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{Q}\mathbf{R}$ where

- $\mathbf{Q}$ is an $m \times n$ matrix with n orthonormal columns: . Construction of $\mathbf{Q}$ is discussed below.

- $\mathbf{R}$ is an $n \times n$ upper triangular matrix: . $\mathbf{R}$ is given by $\mathbf{R} = \mathbf{Q}^{\mathrm{T}}\mathbf{A}$.

This factorisation reduces the normal equation to a much simpler equation:

$$\begin{aligned}
\mathbf{A}^{\mathsf{T}}\mathbf{A}\mathbf{u} &= \mathbf{A}^{\mathsf{T}}\mathbf{b} \\
\Longrightarrow (\mathbf{Q}\mathbf{R})^{\mathsf{T}}\mathbf{Q}\mathbf{R}\mathbf{u}^* &= (\mathbf{Q}\mathbf{R})^{\mathsf{T}}\mathbf{b} \\
\Longrightarrow \mathbf{R}^{\mathsf{T}}\mathbf{Q}^{\mathsf{T}}\mathbf{Q}\mathbf{R}\mathbf{u}^* &= \mathbf{R}^{\mathsf{T}}\mathbf{Q}^{\mathsf{T}}\mathbf{b} \\
\Longrightarrow \mathbf{R}^{\mathsf{T}}\mathbf{R}\mathbf{u}^* &= \mathbf{R}^{\mathsf{T}}\mathbf{Q}^{\mathsf{T}}\mathbf{b} \text{ since } \mathbf{Q}^{\mathsf{T}}\mathbf{Q} = \mathbf{I} \\
\Longrightarrow \mathbf{R}\mathbf{u}^* &= \mathbf{Q}^{\mathsf{T}}\mathbf{b} \text{ multiplying both sides by } (\mathbf{R}^{\mathsf{T}})^{-1}.
\end{aligned}$$

This is easy to solve via back-substitution, since $\mathbf{R}$ is upper triangular.

9.4.1 Constructing the orthogonal matrix Q by Gram-Schmidt

The orthonormal columns of $\mathbf{Q}$, call them $\mathbf{q}_1, \ldots, \mathbf{q}_n$, are obtained iteratively from the columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ of $\mathbf{A}$. The basic idea is that we set $\mathbf{q}_1$ to be $\mathbf{a}_1$. q_2 is then set to be a_2 and any part of it in the direction of $q_1 (= a_1)$ is subtracted out, so ensure that it is orthogonal to q_1. Similarly, q_3 is set to be a_3 with any parts in the direction of q_1 or q_2 are subtracted. All these vectors are normalised to have magnitude 1 at each step.

This is called the *Gram-Schmidt* process and is more formally defined as follows:

$$\begin{array}{lcll}
\text{Set } \mathbf{v}_1 & = & \mathbf{a}_1. & \text{Then } \ \mathbf{q}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|}. \\
\text{Set } \mathbf{v}_2 & = & \mathbf{a}_2 - \left(\mathbf{a}_2^{\mathsf{T}}\mathbf{q}_1\right)\mathbf{q}_1. & \text{Then } \ \mathbf{q}_2 = \frac{\mathbf{v}_2}{|\mathbf{v}_2|}. \\
\vdots & & \vdots & \vdots \qquad \vdots \\
\text{Set } \mathbf{v}_j & = & \mathbf{a}_j - \sum_{i=1}^{j-1}\left(\mathbf{a}_j^{\mathsf{T}}\mathbf{q}_i\right)\mathbf{q}_i. & \text{Then } \ \mathbf{q}_j = \frac{\mathbf{v}_j}{|\mathbf{v}_j|}.
\end{array}$$

Note that $|\mathbf{v}| = \mathbf{v}^{\mathsf{T}}\mathbf{v}$ is the norm of $\mathbf{v}$.

Having found $\mathbf{Q}$, find $\mathbf{R}$ by setting $\mathbf{R} = \mathbf{Q}^{\top}\mathbf{A}$. $\mathbf{R}$ is indeed upper triangular since the ith column of $\mathbf{Q}$ is, by construction, orthogonal to first $i-1$ columns of $\mathbf{A}$.

Producing $\mathbf{Q}$ and $\mathbf{R}$ takes twice as long as the mn^2 steps to form $\mathbf{A}^{\mathsf{T}}\mathbf{A}$, but that extra cost gives a more reliable solution.

There is another method of orthogonalisation that we don't cover here which has better numerical stability using so-called Householder reflectors.

Example: Use the Gram-Schmidt process to orthogonalise the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

Solution: Let $\mathbf{v}_1 = \mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ and then normalise to get $\mathbf{q}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|} = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}$. Now set $\mathbf{v}_2 = \mathbf{a}_2 - (\mathbf{a}_2^\top \mathbf{q}_1)\mathbf{q}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \left(\begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} \right) \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} -0.5 \\ -0.5 \\ 0.5 \\ 0.5 \end{bmatrix}$.

Since $\mathbf{v}_2 = 1$, $\mathbf{q}_2 = \frac{\mathbf{v}_2}{|\mathbf{v}_2|} = \mathbf{v}_2$.

Finally, to get $\mathbf{q}_3$, set

$$\begin{aligned}
\mathbf{v}_3 &= \mathbf{a}_3 - (\mathbf{a}_3^\top \mathbf{q}_1)\mathbf{q}_1 - (\mathbf{a}_3^\top \mathbf{q}_2)\mathbf{q}_2 \\
&= \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} - \left(\begin{bmatrix} 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} \right) \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} - \left(\begin{bmatrix} 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} -0.5 \\ -0.5 \\ 0.5 \\ 0.5 \end{bmatrix} \right) \begin{bmatrix} -0.5 \\ -0.5 \\ 0.5 \\ 0.5 \end{bmatrix} \\
&= \begin{bmatrix} 0.5 \\ -0.5 \\ 0.5 \\ -0.5 \end{bmatrix}
\end{aligned}$$

which is also normalised, so $\mathbf{q}_3 = \mathbf{v}_3$ and

$$\mathbf{Q} = \frac{1}{2} \begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}.$$

We find $\mathbf{R}$ as follows:

$$\mathbf{R} = \mathbf{Q}^\mathsf{T}\mathbf{A} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

□

9.5 Computing $\mathbf{u}^*$ via SVD: the Pseudoinverse

The most stable computation to find the solution to the normal equation is given by singular value decomposition (SVD).
Recall that SVD decomposes the $m \times n$ matrix $\mathbf{A}$ as $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\mathsf{T}$ where:

- $\mathbf{U}$ is a column-orthonormal $n \times m$ matrix, so $\mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{I}_n$,
- $\mathbf{V}$ is an orthonormal $n \times n$ matrix so $\mathbf{V}^\mathsf{T}\mathbf{V} = \mathbf{I}_n$ (indeed, $\mathbf{V}^\mathrm{T} = \mathbf{V}^{-1}$), and
- $\mathbf{D} = \mathrm{diag}\{\sigma_1, \ldots, \sigma_n\}$ is a diagonal $n \times n$ matrix of singular values. Since $\mathbf{D}$ is diagonal, $\mathbf{D}^\mathsf{T} = \mathbf{D}$.

Now consider the product $\mathbf{A}^\mathsf{T}\mathbf{A}$ that arises in the normal equation. Substituting $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\mathsf{T}$ in this product gives:

$$\mathbf{A}^\mathsf{T}\mathbf{A} = (\mathbf{U}\mathbf{D}\mathbf{V}^\mathsf{T})^\mathsf{T}\mathbf{U}\mathbf{D}\mathbf{V}^\mathsf{T} = \mathbf{V}\mathbf{D}^\mathsf{T}\mathbf{U}^\mathsf{T}\mathbf{U}\mathbf{D}\mathbf{V}^\mathsf{T} = \mathbf{V}\mathbf{D}^\mathsf{T}\mathbf{D}\mathbf{V}^\mathsf{T} = \mathbf{V}\mathbf{D}^2\mathbf{V}^\mathsf{T}.$$

We can thus express the normal equation in a much simplified form:

$$\begin{aligned}
\mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{u} &= \mathbf{A}^\mathsf{T}\mathbf{b} \\
\Longrightarrow \mathbf{V}\mathbf{D}^2\mathbf{V}^\mathsf{T}\mathbf{u}^* &= \mathbf{V}\mathbf{D}\mathbf{U}^\mathsf{T}\mathbf{b} \\
\Longrightarrow \mathbf{D}^2\mathbf{V}^\mathsf{T}\mathbf{u}^* &= \mathbf{D}\mathbf{U}^\mathsf{T}\mathbf{b} \\
\Longrightarrow \mathbf{V}^\mathsf{T}\mathbf{u}^* &= \underbrace{\left(\mathbf{D}^2\right)^{-1}\mathbf{D}}_{\mathbf{D}^+}\mathbf{U}^\mathsf{T}\mathbf{b} \\
\Longrightarrow \mathbf{u}^* &= \mathbf{V}\mathbf{D}^+\mathbf{U}^\mathsf{T}\mathbf{b}.
\end{aligned}$$

The matrix $\mathbf{D}^+$ is called the "*pseudoinverse*" of $\mathbf{D}$ and is defined as follows: $\mathbf{D}^+ = \mathrm{diag}\left\{\sigma_1^+, \ldots, \sigma_n^+\right\}$ where

$$\sigma_i^+ = \begin{cases} \sigma_i^{-1} = \frac{1}{\sigma_i} & \text{if} \quad \sigma_i > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Thus if $\mathrm{rank}(\mathbf{A}) = n$, then all the singular values of $\mathbf{A}$ are non-zero, in which case $\mathbf{D}^+ = \mathbf{D}^{-1} = \mathrm{diag}\left\{\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_n}\right\}$. In this case, In the former case, $\mathbf{D}\mathbf{D}^+ = \mathbf{D}^+\mathbf{D} = \mathbf{I}_n$.
However, if $\mathrm{rank}(\mathbf{A}) = r < n$, there are only $r < n$ non-zero singular values and $\mathbf{D}^+ = \mathrm{diag}\{\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_k}, \underbrace{0, \ldots, 0}_{n-r \text{ zeros}}\}$. In this case, $\mathbf{D}\mathbf{D}^+ = \mathbf{D}^+\mathbf{D} = \mathrm{diag}\{1, \ldots, 1, \underbrace{0, \ldots, 0}_{n-r \text{ zeros}}\}$ — which is very close to, but not quite, the identity matrix.
We call the product matrix $\mathbf{V}\mathbf{D}^+\mathbf{U}^\mathsf{T}$ the *pseudoinverse* of $\mathbf{A}$, written $\mathbf{A}^+$. That is ,

$$\mathbf{A}^+ = \mathbf{V}\mathbf{D}^+\mathbf{U}^\mathsf{T}.$$

If $\mathrm{rank}(\mathbf{A}) = n$, then $\mathbf{A}^+\mathbf{A} = \mathbf{V}\mathbf{D}^+\mathbf{U}^\mathsf{T}\mathbf{U}\mathbf{D}\mathbf{V}^\mathsf{T} = \mathbf{V}\mathbf{D}^+\mathbf{D}\mathbf{V}^\mathsf{T} = \mathbf{V}\mathbf{V}^\mathsf{T} = \mathbf{I}_n$, and $\mathbf{A}^+ = \mathbf{A}^{-1}$.

Recall that the matrix $\mathbf{A}^\mathrm{T}\mathbf{A}$ is ill-conditioned when the smallest singular value, σ_n, is very small. This leads to instability in computing the solution to the normal equation. The pseudo-inverse method provides a way of removing this instability to get an approximate but stable solution by simply removing the smallest singular value or values.

9.5.1 Properties of the pseudo inverse $\mathbf{A}^+$

The pseudo-inverse of $\mathbf{A}$ always exists (all we need to do is calculate the SVD and form the product described above).

The SVD of $\mathbf{A}$ gives $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\mathrm{T}$ from which we get $\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{D}$ or, considering the individual columns, $\mathbf{A}\mathbf{v}_i = \sigma_i\mathbf{u}_i$.

- If $\mathbf{A}$ is a square matrix such that $\mathbf{A}^{-1}$ exists, then the singular values for $\mathbf{A}^{-1}$ are $\sigma^{-1} = \frac{1}{\sigma}$ and $\mathbf{A}^{-1}\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{v}_i$
- If $\mathbf{A}^{-1}$ does not exist, then the pseudoinverse matrix $\mathbf{A}^+$ does exist such that:
$$\mathbf{A}^+\mathbf{u}_i = \begin{cases} \frac{1}{\sigma_i}\mathbf{v}_i & \text{if} \quad i \le r = \text{rank}(\mathbf{A}) \text{ i.e. if } \sigma_i > 0 \\ 0 & \text{for} \quad i > r. \end{cases}$$
- Pseudoinverse matrix $\mathbf{A}^+$ has the same rank r as $\mathbf{A}$
- The matrices $\mathbf{A}\mathbf{A}^+$ and $\mathbf{A}^+\mathbf{A}$ are also as near as possible to the $m \times m$ and $n \times n$ identity matrices, respectively
- $\mathbf{A}\mathbf{A}^+$ – the $m \times m$ projection matrix onto the column space of $\mathbf{A}$
- $\mathbf{A}^+\mathbf{A}$ – the $n \times n$ projection matrix onto the row space of $\mathbf{A}$

Example: Find the pseudo-inverse of $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}$.

Solution: The singular value decomposition of $\mathbf{A}$ is

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{bmatrix}}_{\mathbf{U}} \underbrace{\begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{D}} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{\mathbf{V}^\mathrm{T}}.$$

So the pseudo-inverse of $\mathbf{A}$ is

$$\mathbf{A}^+ = \mathbf{V}\mathbf{D}^+\mathbf{U}^\top = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{\mathbf{V}} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{3}} & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{D}^+} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix}}_{\mathbf{U}^\mathrm{T}}$$
$$= \begin{bmatrix} -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \end{bmatrix}.$$

Now let's check the products $\mathbf{AA}^+$ and $\mathbf{A}^+\mathbf{A}$:

$$\mathbf{AA}^+ = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

while

$$\mathbf{A}^+\mathbf{A} = \begin{bmatrix} -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

$\square$

We end our discussion on computational methods in linear algebra here. We have only touched on a small number of the wide variety of techniques used in this area, but we have tried to direct our attention to some of the more common and useful techniques. This is a rich and extremely useful area of study and we encourage interested students to look into fields where these tools and ideas are applied and explored further including computer vision, engineering, physics, applied mathematics and statistics.